TILBURG UNIVERSITY

Tilburg University

The Half-Half Plot

Einmahl, J.H.J.; Gantner, M.

Publication date:
2009

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Einmahl, J. H. J., & Gantner, M. (2009). *The Half-Half Plot*. (CentER Discussion Paper; Vol. 2009-77). Econometrics.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. May. 2023

CentER

Discussion Paper

No. 2009–77

THE HALF-HALF PLOT

By John H.J. Einmahl, Maria Gantner

October 2009

ISSN 0924-7815

Tilburg University

The Half-Half plot*

John H.J. Einmahl Maria Gantner

October 1, 2009

Abstract: The Half-Half (HH) plot is a new graphical method to investigate qualitatively the shape of a regression curve. The empirical HH-plot counts observations in the lower and upper quarter of a strip that moves horizontally over the scatter plot. The plot displays jumps clearly and reveals further features of the regression curve. We prove a functional central limit theorem for the empirical HH-plot, with rate of convergence $1/\sqrt{n}$. In a simulation study the good performance of the plot is demonstrated. The method is also applied to two case studies.

JEL codes: C13, C14.

Key words: Data analysis, functional central limit theorem, graphical methods, jump detection, nonparametric regression.

1 Introduction

Assume the pairs $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$, $n \in \mathbb{N}$, are independent and identically distributed (iid) with bivariate distribution function (df) F. From the regression perspective we can define $\varepsilon = Y - m(X)$ where m, the nonparametric regression function, is a location functional (like the median, mean or mode) applied to the conditional distribution of $Y|X = x$. Equivalently, this can be written in the 'standard form':

$$Y = m(X) + \varepsilon.$$

*John Einmahl (E-mail: *j.h.j.einmahl@uvt.nl*) is Professor of Statistics and Maria Gantner (E-mail: *m.gantner@uvt.nl*) is Ph.D. Candidate of Statistics, Department of Econometrics & OR and CentER, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands.

A first step in data analysis is exploratory diagnostics. Using a good graphical representation of the data sheds light on their main features. The standard procedure for depicting m is kernel estimation, which produces a smooth regression function. When there are sudden changes in m, such as jumps, those are forced into the smoothed picture and therefore ignored. More refined procedures that allow for and detect jumps of m have been introduced and will be discussed later.

In this paper, we present a novel, nonparametric method for explaining regression curves. Its computation is very simple: for small sample sizes, it could even be done by hand. This method displays important features of a regression curve, such as jumps and in- or decreases. If it is a goal to search for jumps, then an ad hoc estimator (introduced in Section 3.2) can be used to find their locations. These procedures impose no particular model on the data; also the regression function m need not be estimated.

Let F be absolutely continuous with density f. Denote the corresponding probability measure with P. Write F_1, F_2 for the marginals of F and Q_1, Q_2 for their (left-continuous) inverse or quantile functions. The empirical counterparts of these functions are denoted with a subscript n, in particular F_n denotes the empirical df of the $(X_i, Y_i), i = 1, \ldots, n$, i.e.

$$F_n(x,y) = \tfrac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]\times(-\infty,y]}(X_i, Y_i), \quad (x,y) \in -\infty < x, y \leq \infty.$$

Fix $\alpha \in (0, \frac{1}{2})$. For $x \in (Q_1(\alpha), Q_1(1-\alpha))$ define the vertical α-strip centered at x by

$$S_\alpha(x) := \{(u,v) \in \mathbb{R}^2 : Q_1(F_1(x) - \alpha) \leq u \leq Q_1(F_1(x) + \alpha)\}.$$

Consider the univariate distribution function on this strip

$$G_{x,\alpha}(y) := \tfrac{1}{2\alpha} \left[F(Q_1(F_1(x)+\alpha), y) - F(Q_1(F_1(x)-\alpha), y)\right], \quad y \in \mathbb{R}.$$

The corresponding quantile function is denoted by $Q_{x,\alpha}$. For convenience, we write in the sequel $x^- = Q_1(F_1(x) - \alpha)$ and $x^+ = Q_1(F_1(x) + \alpha)$.

Definition 1 *Let the coverage $\alpha \in (0, \frac{1}{2})$ be fixed. The Half-Half (HH) value for $x \in (Q_1(\alpha), Q_1(1-\alpha))$ is defined by*

$$\begin{aligned} H_\alpha(x) &= F\left(x, Q_{x,\alpha}\left(\tfrac{1}{4}\right)\right) - F\left(x^-, Q_{x,\alpha}\left(\tfrac{1}{4}\right)\right) \\ &\quad - \left(F\left(x^+, Q_{x,\alpha}\left(\tfrac{3}{4}\right)\right) - F\left(x, Q_{x,\alpha}\left(\tfrac{3}{4}\right)\right)\right) + \tfrac{1}{2}\alpha. \end{aligned}$$

The HH-value is obtained by first vertically dividing the strip $S_\alpha(x)$ into two halves of equal mass: the middle half and the outer half (lower and upper quarter). Then the mass of the lower quarter (i.e. below $Q_{x,\alpha}(\frac{1}{4})$) that is in the left half of the strip (i.e. in $(x^-, x] \times \mathbb{R}$) is added to the mass of the upper quarter that is in the right half of the strip. To standardize the statistic, $\frac{1}{2}\alpha$ is subtracted; this is the sum of these masses corresponding to X and Y being independent. Hence, in that case $H_\alpha \equiv 0$. In general, $H_\alpha \in [-\frac{1}{2}\alpha, \frac{1}{2}\alpha]$. Note that

$$\begin{aligned} H_\alpha(x) &= F\left(x, Q_{x,\alpha}\left(\tfrac{1}{4}\right)\right) - F\left(x^-, Q_{x,\alpha}\left(\tfrac{1}{4}\right)\right) \\ &\quad + F\left(x, Q_{x,\alpha}\left(\tfrac{3}{4}\right)\right) - F\left(x^-, Q_{x,\alpha}\left(\tfrac{3}{4}\right)\right) - \alpha. \end{aligned}$$

Definition 2 *Let $\alpha \in (0, \frac{1}{2})$. The Half-Half (HH) plot is the graph of the function*

$$x \mapsto H_\alpha(x),\ x \in (Q_1(\alpha), Q_1(1-\alpha))\,.$$

Define for $x \in (Q_{n,1}(\alpha), Q_{n,1}(1-\alpha))$ the empirical counterpart of $G_{x,\alpha}$ by

$$G_{n,x,\alpha}(y) = \tfrac{1}{2\alpha}\left[F_n(Q_{n,1}(F_{n,1}(x)+\alpha), y) - F_n(Q_{n,1}(F_{n,1}(x) - \tfrac{\lceil n\alpha\rceil}{n}), y)\right], \quad y \in \mathbb{R},$$

where $\lceil\cdot\rceil$ denotes the ceiling function. Write $X_{[1]} \le X_{[2]} \le \cdots \le X_{[n]}$ for the order statistics of the $X_i, i = 1, \ldots, n$. It is easily seen that

$$G_{n,x,\alpha}(y) = \tfrac{1}{2\alpha}\left[F_n(X_{[i+\lceil n\alpha\rceil]}, y) - F_n(X_{[i-\lceil n\alpha\rceil]}, y)\right],$$

where i is such that $X_{[i]} \le x < X_{[i+1]}$. Denote the quantile function of $G_{n,x,\alpha}$ with $Q_{n,x,\alpha}$. The HH-statistic is for $x \in (Q_{n,1}(\alpha), Q_{n,1}(1-\alpha))$ defined to be the empirical version of the HH-value:

$$\begin{aligned} H_{n,\alpha}(x) &= F_n\left(x, Q_{n,x,\alpha}\left(\tfrac{1}{4}\right)\right) - F_n\left(Q_{n,1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha\rceil}{n}\right), Q_{n,x,\alpha}\left(\tfrac{1}{4}\right)\right) \\ &\quad + F_n\left(x, Q_{n,x,\alpha}\left(\tfrac{3}{4}\right)\right) - F_n\left(Q_{n,1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha\rceil}{n}\right), Q_{n,x,\alpha}\left(\tfrac{3}{4}\right)\right) - \alpha \\ &= F_n\left(X_{[i]}, Q_{n,x,\alpha}\left(\tfrac{1}{4}\right)\right) - F_n\left(X_{[i-\lceil n\alpha\rceil]}, Q_{n,x,\alpha}\left(\tfrac{1}{4}\right)\right) \\ &\quad + F_n\left(X_{[i]}, Q_{n,x,\alpha}\left(\tfrac{3}{4}\right)\right) - F_n\left(X_{[i-\lceil n\alpha\rceil]}, Q_{n,x,\alpha}\left(\tfrac{3}{4}\right)\right) - \alpha. \end{aligned}$$

The empirical HH-plot is defined to be the graph of the function

$$x \mapsto H_{n,\alpha}(x), \quad x \in (Q_{n,1}(\alpha), Q_{n,1}(1-\alpha))\,.$$

Jumps of m are depicted in the HH-plot by a high positive (jump up) or negative (jump down) value. In the first example, we consider $m_1(x) = \mathbb{1}_{[0.5,\infty)}(x)$, with X uniformly

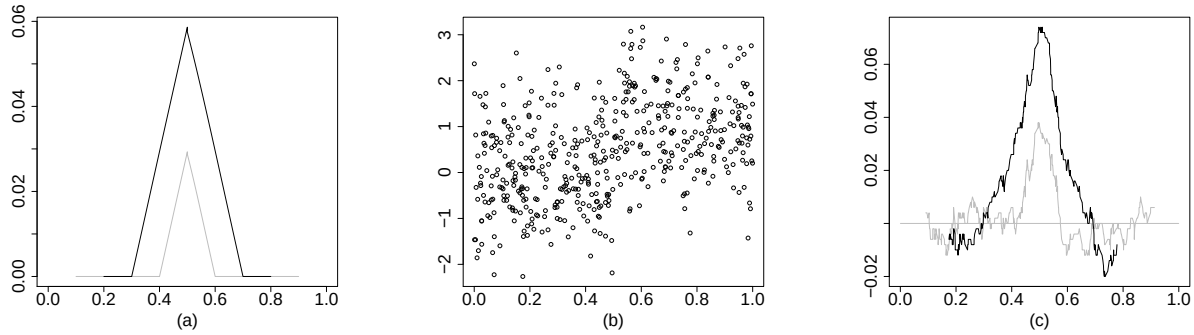

Figure 1: Theoretical (a) and empirical (c) HH-plot with $\alpha = 0.1$ (grey line) and $\alpha = 0.2$ (black line) of a sample of size $n = 500$ (b) of m_1 with ε standard normal and $X \sim UN(0, 1)$.

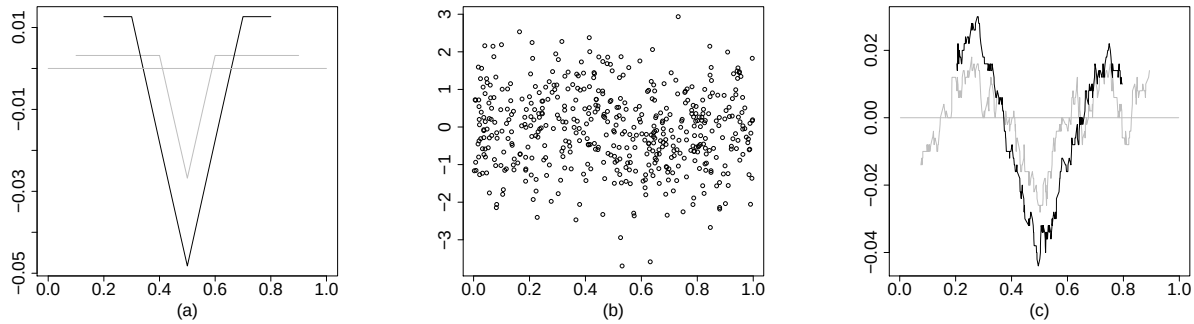

Figure 2: Theoretical (a) and empirical (c) HH-plot with $\alpha = 0.1$ (grey line) and $\alpha = 0.2$ (black line) of a sample of size $n = 500$ (b) of m_2 with ε standard normal and $X \sim UN(0, 1)$.

distributed on $(0, 1)$ and ε standard normal and independent of X, see Figure 1. The jump is clearly indicated by the large HH-value and -statistic at $x = 0.5$. Additionally, the HH-plot reveals that the regression curve is constant before and after the jump, which is indicated by values of $H_\alpha(x)$ and $H_{n,\alpha}(x)$ close to 0.

Another important feature of the HH-plot can be observed in Figure 2. Let X and ε be as above, but $m_2(x) = x - \mathbb{1}_{[0.5,\infty)}(x)$. The jump down is indicated by large negative values of $H_\alpha(0.5)$ and $H_{n,\alpha}(0.5)$, whereas the positive values indicate that m_2 is increasing. Hence the HH-plot can be used for detecting jumps as well as continuous increases or decreases of the regression curve.

It is advised to calculate the HH-plot for two values of α. Changes in the regression are typically easier detected by using larger α-strips. But because the empirical HH-statistic is only defined for $x \in (Q_{n,1}(\alpha), Q_{n,1}(1-\alpha))$, a larger α reduces the range. A natural

maximum here seems to be $\alpha = 0.25$. Hence, in order to get an impression of how the regression curve behaves in the beginning and the end, additionally a HH-plot with a small α should be depicted. A rule of thumb is that for $n \geq 200$ the larger α should be taken 0.2, and the smaller α can be chosen such that $n\alpha \geq 25$; for $n < 200$ we advise to draw the HH-plot for only one α, for instance 0.2.

There exists a large body of literature on the specific topic of estimating jump points in nonparametric regression. A good overview can, for example, be found in Gijbels et al. (2007). The most common approach is to compare left- and right-sided estimators of the regression function at point x. These estimators are attained by kernel methods [amongst others, Müller (1992), Hall and Titterington (1992), and Wu and Chu (1993)] or local polynomial regression [amongst others, McDonald and Owen (1986), Loader (1996), Horváth and Kokoszka (2002), Grégoire and Hamrouni (2002), and Gijbels et al. (2007)]. Some papers propose a two-step procedure, where in a first step the location of jump is estimated and in a second step the curve is fitted [cf. Gijbels et al. (1999), Sánchez-Borrego et al. (2006)]. Wavelet methods are, amongst others, employed in Wang (1995), Antonidiadis and Gijbels (2002) and Park and Kim (2006). In Kim and Marron (2006), another graphical method, the SiZer for jump detection, is introduced. The SiZer is a kernel-based approach which combines multiple bandwidths in one plot. Its main goal is to get a first, general overview of the regression curve.

In the papers mentioned above, the specific assumptions on the regression curve and the errors play an important role. In contrast, we have no direct assumptions on m or ε. In Dempfle and Stute (2002) an approach based on U-statistics is used. As in the present paper, this approach requires minimal assumptions on the model and avoids smoothing for estimating the location of the jump.

The paper is organized as follows. The asymptotic behavior of the HH-statistic is expounded in Section 2, where it is shown that the rate of convergence of the empirical HH-statistic to the theoretical one is $1/\sqrt{n}$, uniformly in x. The limiting process of the HH-value is also used to facilitate the depiction of jumps. In Section 3 three simulation studies are presented and an ad hoc estimator for a jump location is introduced in Section 3.2. Two real data applications can be found in Section 4. The paper is completed by a section containing the proof of the result of Section 2.

2 Asymptotic results

Define for $n \in \mathbb{N}$ the empirical process indexed by points as

$$U_n(x,y) := n^{\frac{1}{2}}\big\{F_n(x,y) - F(x,y)\big\}, \quad -\infty < x, y \leq \infty.$$

Furthermore, let B_F be a bounded, mean zero Gaussian process on $(-\infty,\infty]^2$ that is uniformly continuous and has covariance function $F(x_1 \wedge x_2, y_1 \wedge y_2) - F(x_1,y_1)F(x_2,y_2)$, $(x_1,y_1),(x_2,y_2) \in (-\infty,\infty]^2$. Then, by the functional central limit theorem for U_n and the Skorohod representation theorem, there exist $\tilde{B}_F \stackrel{d}{=} B_F$ and a sequence $\tilde{U}_n \stackrel{d}{=} U_n$, $n \in \mathbb{N}$, such that

$$\sup_{-\infty<x,y\leq\infty} \big|\tilde{U}_n(x,y) - \tilde{B}_F(x,y)\big| \to 0 \quad \text{a.s. as } n \to \infty. \tag{1}$$

Henceforth we will use (1), but drop the tildes from the notation.

We will use the following assumptions:

(A) Let $\mathcal{S} := \{(x,y) \in \mathbb{R}^2 : f(x,y) > 0\} = (x_*, x^*) \times \mathbb{R}$ for some $-\infty \leq x_* < x^* \leq \infty$ and F be of class C^2 on $\mathcal{S}$ and f_2 be bounded.

Hence, the function $G_{x,\alpha}(\cdot)$ is increasing and continuous on $(Q_1(\alpha), Q_1(1-\alpha))$; also we can write $F'_x(x,y) := \frac{\partial}{\partial x}F(x,y) = \int_{-\infty}^{y} f(x,v)dv$ and $F'_y(x,y) := \frac{\partial}{\partial y}F(x,y) = \int_{-\infty}^{x} f(u,y)du$.

For the second assumption we first introduce some more notation: Let $\tilde{F}(u,y) := F(Q_1(u),y)$ denote the in the first coordinate uniformized distribution function, and write $\tilde{F}'_x(u,y) := \frac{\partial}{\partial u}\tilde{F}(u,y) = \frac{F'_x(Q_1(u),y)}{f_1(Q_1(u))}$ and $\tilde{F}'_y(u,y) := \frac{\partial}{\partial y}\tilde{F}(u,y) = F'_y(Q_1(u),y)$.

(B) $\tilde{F}'_x$ and $\tilde{F}'_y$ are uniformly continuous on $(0,1) \times \mathbb{R}$.

Let $\alpha \in (0, \frac{1}{2})$ and define $I_n := (Q_1(\alpha), Q_1(1-\alpha)) \cap (Q_{n,1}(\alpha), Q_{n,1}(1-\alpha))$ and $I_0 := (Q_1(\alpha), Q_1(1-\alpha))$. Denote the HH-process with

$$V_{n,\alpha}(x) := n^{\frac{1}{2}}\big(H_{n,\alpha}(x) - H_\alpha(x)\big), \quad x \in I_n,$$

and for $s \in \{\frac{1}{4}, \frac{3}{4}\}$ define

$$
\begin{aligned}
B_{\alpha,s}(x) \;=\;& B_F\big(x, Q_{x,\alpha}(s)\big) \;-\; B_F\big(x^-, Q_{x,\alpha}(s)\big) \\
&- \frac{F'_x\big(x^-, Q_{x,\alpha}(s)\big)}{f_1\big(x^-\big)}\big[B_F\big(x,\infty\big) \;-\; B_F\big(x^-,\infty\big)\big] \\
&+ \frac{F'_y\big(x^-, Q_{x,\alpha}(s)\big) \;-\; F'_y\big(x, Q_{x,\alpha}(s)\big)}{F'_y\big(x^+, Q_{x,\alpha}(s)\big) - F'_y\big(x^-, Q_{x,\alpha}(s)\big)}\Big\{B_F\big(x^+, Q_{x,\alpha}(s)\big) - B_F\big(x^-, Q_{x,\alpha}(s)\big) \\
&\quad + \frac{F'_x\big(x^+, Q_{x,\alpha}(s)\big)}{f_1\big(x^+\big)}\;\big[B_F(x,\infty) \;-\; B_F\big(x^+,\infty\big)\big] \\
&\quad - \frac{F'_x\big(x^-, Q_{x,\alpha}(s)\big)}{f_1\big(x^-\big)}\;\big[B_F(x,\infty) \;-\; B_F\big(x^-,\infty\big)\big]\Big\}, \quad x \in I_0.
\end{aligned} \tag{2}
$$

Theorem 1 *Under the assumptions (A) and (B), we have for all $\alpha \in (0, \frac{1}{2})$, on the probability space of (1),*

$$\sup_{x\in I_n} \big|V_{n,\alpha}(x) - \big(B_{\alpha,1/4}(x) + B_{\alpha,3/4}(x)\big)\big| \to 0 \quad a.s. \quad as \;\; n \to \infty.$$

Remark 1 *For X and Y independent (m is constant), the distribution of the limiting process from Theorem 1 for fixed $x \in I_0$ simplifies to a $N(0, \frac{1}{4}\alpha)$-distribution.*

Proof of Remark 1 In this case the limiting process specializes to

$$
\begin{aligned}
B_{\alpha,1/4}(x) + B_{\alpha,3/4}(x) \;=\;& \tfrac{1}{2}\Big\{\big[B_F(x, Q_2\left(\tfrac{1}{4}\right)) - B_F(x^-, Q_2\left(\tfrac{1}{4}\right))\big] \\
&- \big[B_F(x^+, Q_2\left(\tfrac{1}{4}\right)) - B_F(x, Q_2\left(\tfrac{1}{4}\right))\big] \;+\; \big[B_F(x, Q_2\left(\tfrac{3}{4}\right)) - B_F(x^-, Q_2\left(\tfrac{3}{4}\right))\big] \\
&- \big[B_F(x^+, Q_2\left(\tfrac{3}{4}\right)) - B_F(x, Q_2\left(\tfrac{3}{4}\right))\big] \;-\; \big[B_F(x,\infty) - B_F(x^-,\infty)\big] \\
&+ \big[B_F(x^+,\infty) - B_F(x,\infty)\big]\Big\}.
\end{aligned}
$$

Taking the sets $A_s := (x^-, x] \times (-\infty, Q_2(s)]$, $B_s := (x, x^+] \times (-\infty, Q_2(s)]$, $C_s := (x^-, x] \times (Q_2(s), \infty)$ and $D_s := (x, x^+] \times (Q_2(s), \infty)$, as depicted in Figure 3, as well as $E^- := (x^-, x] \times \mathbb{R}$ and $E^+ := (x, x^+] \times \mathbb{R}$, into account, we have $P(A_s) = P(B_s) = s\alpha$, $P(C_s) = P(D_s) = (1-s)\alpha$, $P(E^-) = P(E^+) = \alpha$.

Extending the definition of B_F to semi-infinite rectangles in the usual way, we get

$$
\begin{aligned}
&B_{\alpha,1/4}(x) + B_{\alpha,3/4}(x) \\
&= \tfrac{1}{2}\big\{B_F(A_{1/4}) - B_F(B_{1/4}) + B_F(A_{3/4}) - B_F(B_{3/4}) - B_F(E^-) + B_F(E^+)\big\} \\
&= \tfrac{1}{2}\big\{B_F(A_{1/4}) - B_F(B_{1/4}) - B_F(C_{3/4}) + B_F(D_{3/4})\big\}.
\end{aligned} \tag{3}
$$

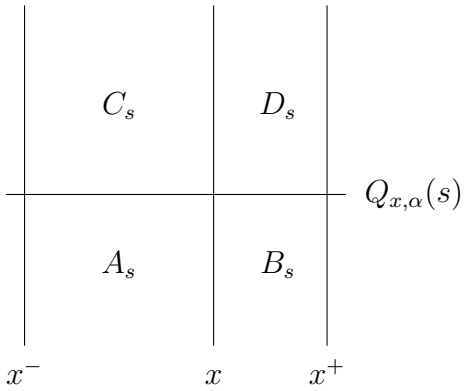

Figure 3: Regions taken into account for the calculation of the limiting process at x, where $s \in \{\frac{1}{4}, \frac{3}{4}\}$.

Let W_F be the Wiener process defined on semi-infinite rectangles R by $W_F(R) = B_F(R) + P(R)Z$, where $Z \sim N(0,1)$ and B_F are independent. Then the right-hand side of (3) can be rewritten as

$$\frac{1}{2}\left\{W_F(A_{1/4}) - W_F(B_{1/4}) - W_F(C_{3/4}) + W_F(D_{3/4})\right\}, \tag{4}$$

which is easily seen to be $N(0, \frac{1}{4}\alpha)$-distributed. □

3 Simulation study

For an easier interpretation of the HH-plot, we add a horizontal band to the picture. If the empirical HH-plot escapes the band, this indicates jumps or steep in- or decreases of the regression curve. In addition, this band gives a standard to assess the relative magnitude of the HH-statistic. The band is obtained by calculating high quantiles of

α	0.05	0.1	0.15	0.2	0.25
$q_\alpha(0.9)$	0.40	0.53	0.61	0.67	0.72
$q_\alpha(0.95)$	0.42	0.56	0.66	0.73	0.79
$q_\alpha(0.99)$	0.47	0.64	0.75	0.84	0.91

Table 1: The 0.90, 0.95 and 0.99 quantiles of the random variable in (5), for $\alpha \in \{0.05, 0.1, 0.15, 0.2, 0.25\}$.

(5) $$\sup_{x\in I_0} \left| B_{\alpha,1/4}(x) + B_{\alpha,3/4}(x) \right|,$$

in case X and Y are independent; (4) is useful for this calculation. These quantiles are denoted q_α (see Table 1) and have to be divided by $\sqrt{n}$ and $-\sqrt{n}$, respectively, in order to obtain the upper and lower boundary of the band.

3.1 X and Y independent

First, we consider the case of X,Y being independent, with sample sizes of $n = 250$ and 500 for coverage levels $\alpha = 0.1$ and 0.2. For each n, 10,000 samples are taken. Table 2 provides the fraction of exceedances of $q_\alpha(0.95)/\sqrt{n}$ by $\max\{|H_{n,\alpha}(x)| : Q_{n,1}(\alpha) < x < Q_{n,1}(1-\alpha)\}$. We see that these numbers are close to, but somewhat lower than, $1 - 0.95 = 0.05$. This might be due to the fact that for fixed x, the effective sample size (i.e. the number of observations in the lower and upper quarter of the strip) is $n\alpha$, which can be as small as 25. Indeed, for $n = 500$ and $\alpha = 0.2$, we see that the asymptotic quantiles are quite accurate.

	$\max(\lvert H_{n,\alpha}(x)\rvert) \geq q_\alpha(0.95)/\sqrt{n}$	
n	$\alpha = 0.1$	$\alpha = 0.2$
250	0.025	0.032
500	0.037	0.044

Table 2: Simulated Type I error probabilities for X and Y independent.

3.2 Regression functions m_1 and m_2 of Section 1

Next, we consider m_1 and m_2 as in Section 1. We simulate from the models presented there, and additionally take the parameters $\sigma = 0.1$ and $\sigma = 0.5$ for the normal distribution of the errors into account. Sample sizes and numbers of replications are taken as above. As before, the upwards (m_1) and downwards (m_2) fraction of exceedances of $q_\alpha/\sqrt{n}$ and $-q_\alpha/\sqrt{n}$, respectively, is measured. In addition, the fraction of how often there is indeed a jump up of m_1 detected at $x = 0.5$, thus how often the empirical HH-plot at $x = 0.5$ exceeds $q_\alpha(0.95)/\sqrt{n}$, is reported. Similarly, this fraction is also given for the jump down of m_2.

m_1		$\max(H_{n,\alpha}(x)) \geq q_\alpha(0.95)/\sqrt{n}$		$H_{n,\alpha}(0.5) \geq q_\alpha(0.95)/\sqrt{n}$		$\mathrm{sd}(\hat{\theta})$	
σ	n	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$
0.1	250	1	1	1	1	0.007	0.007
	500	1	1	1	1	0.003	0.003
0.5	250	0.993	1	0.979	1	0.011	0.011
	500	1	1	1	1	0.006	0.006
1	250	0.412	0.957	0.243	0.858	0.069	0.038
	500	0.902	1	0.766	0.998	0.027	0.020

Table 3: Simulation study as described in Section 3.2 for the regression function m_1.

m_2		$\min(H_{n,\alpha}(x)) \leq -q_\alpha(0.95)/\sqrt{n}$		$H_{n,\alpha}(0.5) \leq -q_\alpha(0.95)/\sqrt{n}$		$\mathrm{sd}(\hat{\theta})$	
σ	n	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$
0.1	250	1	1	1	1	0.005	0.004
	500	1	1	1	1	0.002	0.002
0.5	250	0.965	1	0.931	1	0.012	0.012
	500	1	1	1	1	0.006	0.006
1	250	0.289	0.737	0.159	0.569	0.070	0.038
	500	0.777	0.985	0.616	0.958	0.026	0.020

Table 4: Simulation study as described in Section 3.2 for the regression function m_2.

The exceedance rates close to 100% in Tables 3 and 4 show that the HH-plot is a very good tool to detect jumps. Only at $\sigma = 1$, when the picture is very blurry, the effective sample size of $n\alpha = 25$ ($n = 250$, $\alpha = 0.1$) is too small and the HH-plot has less power.

Last, an ad hoc estimator $\hat{\theta}$ of the jump location is given, by taking in every simulation the mean of $\mathrm{argmax}\{H_{n,\alpha}(x) : x \in (Q_{n,1}(\alpha), Q_{n,1}(1-\alpha))\}$ (and, for m_2, similarly the mean of the argmin). Due to symmetry, this estimator is unbiased. The standard deviations, presented in the last two columns of Tables 3 and 4, show that this approach of averaging the locations of the extrema leads to a good estimator of the jump location.

3.3 Regression function as in Gijbels et al. (1999)

The final simulation is an example from Gijbels et al. (1999), with regression function

$$m_3(x) = \begin{cases} \exp\{-2(x-0.35)\} - 1 & \text{if} \quad x \in [0, 0.35) \\ \exp\{-2(x-0.35)\} & \text{if} \quad x \in [0.35, 0.65) \\ \exp\{2(x-0.65)\} + \exp\{-0.6\} - 2 & \text{if} \quad x \in [0.65, 1]. \end{cases} \tag{6}$$

It is depicted in Figure 4(a); a scatter plot with $X \sim UN(0,1)$, ε standard normal and independent of X, and $n = 250$ is given in Figure 4(b).

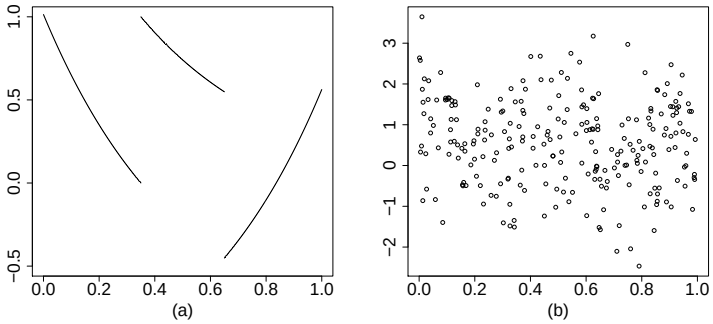

Figure 4: Panel (a) shows m_3 as given in (6) and panel (b) one simulation of m_3 with ε standard normal, $X \sim UN(0,1)$ and $n = 250$.

In Figure 5, the theoretical HH-plots for m_3 with $X \sim UN(0,1)$ and $\varepsilon \sim N(0, \sigma^2)$, $\sigma \in \{0.1, 0.5, 1\}$, for both $\alpha = 0.1$ and $\alpha = 0.2$, are given. The features of the HH-plot are nicely shown in Figure 5(a). The possible range of $H_{0.1}$ is $[-0.05, 0.05]$, and that of $H_{0.2}$ is $[-0.10, 0.10]$, and in both cases it is fully utilized. The negative slope of the first part of m_3 is depicted in the negative, almost horizontal line of the HH-plot for $\alpha = 0.1$ and the negative values of $H_{0.2}$. The latter one starts at $x = 0.2$ and hence already takes the jump upwards into account ($2\alpha = 0.4 > 0.35$), which is depicted by the increasing HH-values. For $\alpha = 0.1$, this increase starts at $x = 0.35 - \alpha = 0.25$. The jump point at $x = 0.35$ is for both coverage levels α depicted by a distinct local maximum. The regression curve after this jump is again decreasing, hence the HH-values become negative, even before $x^- > 0.35$. For $x \in (0.45, 0.55)$, the plot of $H_{0.1}$ only depicts the negative slope, and hence almost reaches its minimal possible value. The jump at $x = 0.65$ is then difficult to see from the HH-plot, because the HH-values for both $\alpha = 0.1$ and $\alpha = 0.2$ only change slightly in a region left of the jump point. The last part of m_3 has a positive slope, hence the HH-values become positive again.

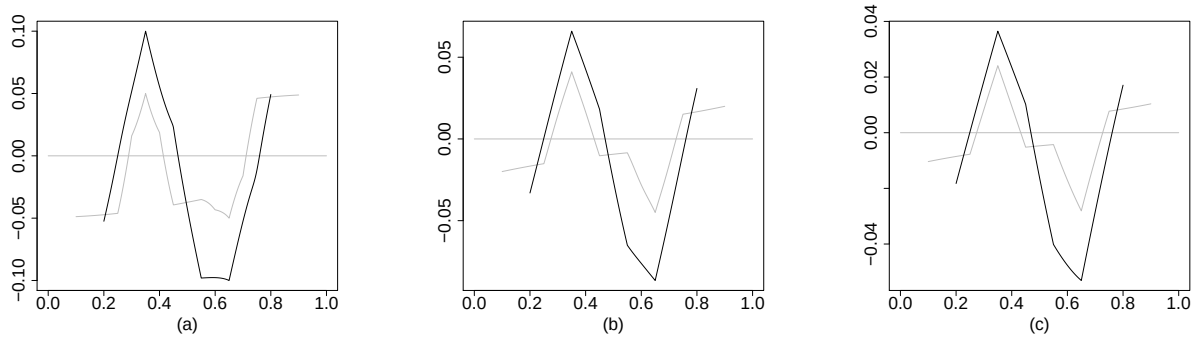

Figure 5: The theoretical HH-plot for m_3 is shown for various standard deviations ((a) $\sigma = 0.1$, (b) $\sigma = 0.5$ and (c) $\sigma = 1$) of the normal error, for $\alpha = 0.1$ (grey line) and $\alpha = 0.2$ (black line).

Figures 5(b) and (c) bear similar features as (a). The jumps are indicated by the extrema of the HH-plot, the steep in- and decreases by positive/negative HH-values. But in contrast to panel (a), the jumps down of the regression curve are depicted by distinct minima of the HH-plot. Also observe that the range of the HH-values is getting smaller when σ increases.

In Table 5, the outcomes of the simulations are presented, with columns as described for m_1 and m_2. The standard deviations of $\hat{\theta}$ are moved to Table 6. The outcomes of the simulations presented in Table 5 are very similar to the outcomes of the simulations for m_1 and m_2; in general the HH-plot depicts the main features of the regression curve very well, but again the combination of a smaller $n\alpha$ with a larger σ results in a lower power.

m_3		$\max(H_{n,\alpha}(x)) \geq q_\alpha(0.95)/\sqrt{n}$		$H_{n,\alpha}(0.35) \geq q_\alpha(0.95)/\sqrt{n}$		$\min(H_{n,\alpha}(x)) \leq -q_\alpha(0.95)/\sqrt{n}$		$H_{n,\alpha}(0.65) \leq -q_\alpha(0.95)/\sqrt{n}$	
σ	n	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$
0.1	250	1	1	1	1	1	1	1	1
	500	1	1	1	1	1	1	1	1
0.5	250	0.911	0.978	0.821	0.949	0.991	1	0.960	1
	500	1	1	0.999	1	1	1	1	1
1	250	0.210	0.378	0.091	0.228	0.389	0.911	0.198	0.726
	500	0.658	0.794	0.444	0.658	0.882	0.999	0.695	0.989

Table 5: Simulation study as described in Section 3.3 for the regression function m_3.

In Table 6 the biases and standard deviations of the extrema of the HH-statistic are presented. [Here we have restricted argmax $H_{n,\alpha}$ in the definition of $\hat{\theta}_{max}$ to $(Q_{n,1}(\alpha), \frac{1}{2}]$ (and similarly argmin $H_{n,\alpha}$ to $[\frac{1}{2}, Q_{n,1}(1-\alpha))$) in order not to confuse the steep increase for large x with the jump at $x = 0.35$.] The estimator of the jump up (at $x = 0.35$) is indicated as $\hat{\theta}_{max}$, that of the jump down (at $x = 0.65$) with $\hat{\theta}_{min}$. For comparison with Gijbels et al. (1999), we also include $n = 100$. As expected from the above discussion of Figure 5, $\hat{\theta}_{min}$ has a negative bias and a larger standard deviation. It is remarkable, but explainable through Figure 5, that $\hat{\theta}_{min}$ is performing better in the more 'difficult' setup of $\sigma = 0.5$ than in that of $\sigma = 0.1$.

Although estimation is not our main goal, our estimator compares well with Gijbels et al. (1999). We liken our best results to the best results of Gijbels et al. (1999). The estimators for the jump locations there will be denoted in the following with $\hat{\theta}^G_{max}$ for the first jump at $x = 0.35$, and $\hat{\theta}^G_{min}$ for the second jump at $x = 0.65$. The absolute values of the biases of our $\hat{\theta}_{max}$ are always smaller or equal than those of $\hat{\theta}^G_{max}$; the standard deviations are similar. For the jump down at $x = 0.65$, σ plays a role. For $\sigma = 0.1$, the absolute values of the biases and the standard deviations of $\hat{\theta}_{min}$ are both greater or equal than those of $\hat{\theta}^G_{min}$. But with σ increased, $\hat{\theta}_{min}$ has similar biases and even smaller standard deviations than $\hat{\theta}^G_{min}$. To conclude, our ad hoc estimator is a competing method to detect jump locations, especially recommended when the picture is blurry.

m_3		bias($\hat{\theta}_{max}$)		sd($\hat{\theta}_{max}$)		bias($\hat{\theta}_{min}$)		sd($\hat{\theta}_{min}$)	
σ	n	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$
	100	0.0004	0.0003	0.009	0.008	−0.042	−0.049	0.025	0.020
0.1	250	0.0000	0.0001	0.004	0.003	−0.019	−0.043	0.022	0.020
	500	0.0000	0.0000	0.002	0.001	−0.006	−0.034	0.011	0.022
	100	0.0019	0.0060	0.025	0.026	−0.012	−0.022	0.028	0.029
0.5	250	0.0007	0.0027	0.011	0.012	−0.005	−0.015	0.013	0.020
	500	0.0002	0.0016	0.006	0.007	−0.003	−0.009	0.007	0.014
	100	−0.0043	0.0067	0.059	0.049	−0.012	−0.029	0.054	0.045
1	250	0.0019	0.0071	0.030	0.031	−0.012	−0.027	0.033	0.034
	500	0.0011	0.0047	0.017	0.019	−0.009	−0.024	0.021	0.029

Table 6: Simulation study as described in Section 3.3 for the regression function m_3.

4 Real data applications

The real data examples in this section are based on a fixed, equidistant design, rather than a random regressor X. In this case we observe

$$Y_{n,i} = m\left(\tfrac{i}{n+1}\right) + \varepsilon_{n,i}, \quad \text{with independent and centered } \varepsilon_{n,i} \sim F_{n,i}\,, i = 1, \ldots, n.$$

If all the $F_{n,i}$, $i = 1, \ldots, n$, are equal and sufficiently smooth, then the limiting process is as in Remark 1 and (5), therefore Table 1 remains applicable, for n large enough. In the finite sample case, the only minor difference between fixed and random design is the domain of $V_{n,\alpha}$, which is I_0 in the above fixed design setting.

4.1 Nile data

The Nile data, reported in Cobb (1978), are a popular dataset first used in the context of nonparametric change-point estimation in Carlstein (1988). Figure 6, left panel, shows the annual volume of discharge ($10^{10}m^3$) from the Nile River for each year from 1871 through 1970. In Cobb (1978) it is assumed that the observations are independent, and parametric methods are used to find a change-point in 1897. Meteorological studies confirm this change. According to Kraus (1955), the rainfall decreased in most regions abruptly at the

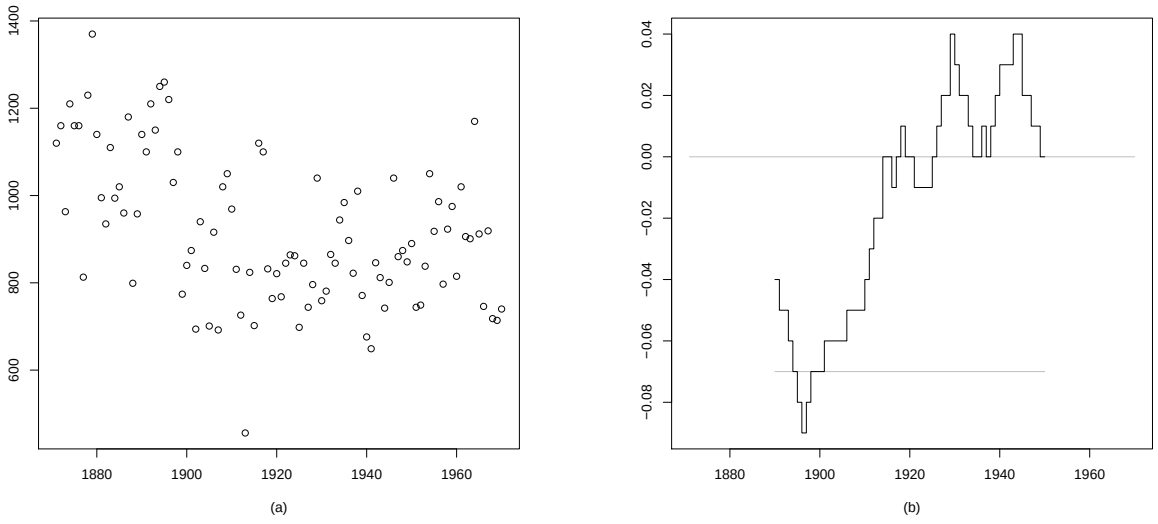

Figure 6: Annual volume discharge of the Nile; (a) scatter plot, (b) HH-plot for $\alpha = 0.2$.

end of the 19th century, which was due to a narrowing of the rainfall belt and a shortening of the wet seasons.

The HH-plot with $\alpha = 0.2$ displays this change clearly. Because of the small sample size, we calculate the horizontal band directly from a simulation for equal $F_{n,i},\ i = 1, \ldots, n;\ n = 100$, cf. the beginning of Section 3. This gives a value of ± 0.07 for the boundary of the band. Because the HH-statistics are below the band in 1895–1897, we can conclude that there is an abrupt change of the regression curve which is located, according to the ad hoc method, at 1896. In the second half of the period under consideration, the regression curve seems to be slightly increasing. Here, the HH-plot performs well for a sample of relatively small size.

4.2 Prague temperatures

As a second application, we consider the average annual temperatures in Prague from 1775 through 1989. In previous analyses it is found that the number of jump points is two or three, see Table 7. (Note that in Antonidiadis and Gijbels (2002) the data are studied only up to 1902.) From the HH-plot, the change-point recurring in the literature at around 1835 is easily detected. The plot, see Figure 7(b), is for the years in 1823–1835 below

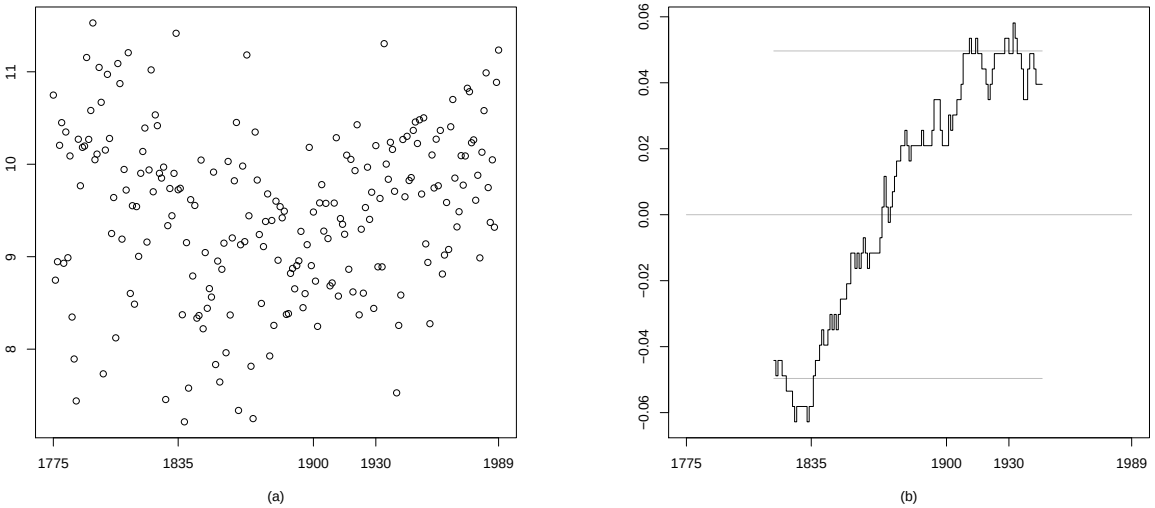

Figure 7: Average annual temperatures in Prague from 1775 to 1989; (a) scatter plot, (b) HH-plot for $\alpha = 0.2$.

methods	years of jump			
Horváth, Kokoszka and Steinebach (1999)		1835	1893	1927
Antonidiadis and Gijbels (2002)	1787	1837		–
Gijbels and Goderniaux (2004)	1786.5	1836.5		1942.5
HH-plot		1830		1932

Table 7: Jump points in the literature.

$-q_{0.2}(0.95)/\sqrt{n}$ and in 1911, 1914, 1928–1929 and 1932–1933 above $q_{0.2}(0.95)/\sqrt{n}$. Hence there is a drastic change between 1823 and 1835; according to the spikes we might locate it at 1830. After this jump down follows a strong increase of the regression curve with a possible jump up between 1911 and 1933. Using the ad hoc estimator, we can locate such a jump in the year 1932. This change is in accordance with changes found in the literature (1927 and 1942.5). The early jump found in the literature (at 1787) is difficult to assess since there are only 12 observations before 1787; this year is outside the domain of the HH-plot.

5 Proofs

Because α is fixed, it is henceforth dropped from the notation as a subscript. Set $\tilde{F}_n(u,y) := F_n(Q_1(u),y)$, $\tilde{F}_{n,1} := \tilde{F}_n(u,\infty)$, and denote its generalized inverse by $\tilde{F}_{n,1}^{-1}$. The corresponding standard Brownian bridge is $B_1(u) := B_F(Q_1(u),\infty)$. Because the marginals of $\tilde{F}(u,y)$, the identity and F_2, are uniformly continuous, $\tilde{F}$ itself is also uniformly continuous on $(0,1)\times\mathbb{R}$, and for all $(u,y)\in(0,1)\times\mathbb{R}$, since $\tilde{F}(u,\infty)=u$,

$$\tilde{F}'_x(u,y) \le 1. \tag{7}$$

5.1 Lemmas

The proof of Theorem 1 is based on three lemmas.

Lemma 1 *Let* $S_n(u,y) := n^{\frac{1}{2}}\left(\tilde{F}\left(\tilde{F}_{n,1}^{-1}(u),y\right)-\tilde{F}\left(u,y\right)\right)$, *with* $0<u<1$, $y\in\mathbb{R}$, $n\in\mathbb{N}$. *With assumptions (A) and (B), on the probability space of (1),*

$$\sup_{0<u<1}\sup_{y\in\mathbb{R}}\left|S_n(u,y)+\tilde{F}'_x(u,y)B_1(u)\right|\to 0 \quad a.s.\ \ as\ \ n\to\infty. \tag{8}$$

Proof With the mean-value theorem with u^* between u and $\tilde{F}_{n,1}^{-1}(u)$, we can write $S_n(u,y) = n^{\frac{1}{2}}\left(\tilde{F}_{n,1}^{-1}(u) - u\right)\tilde{F}'_x(u^*,y)$, and hence the left-hand side of (8) is bounded from above by

$$\begin{aligned} &\sup_{0<u<1}\left|n^{\frac{1}{2}}\left(\tilde{F}_{n,1}^{-1}(u) - u\right) + B_1(u)\right| \sup_{0<u<1}\sup_{y\in\mathbb{R}}\tilde{F}'_x(u^*,y) \\ &+ \sup_{0<u<1}|B_1(u)| \sup_{0<u<1}\sup_{y\in\mathbb{R}}\left|\tilde{F}'_x(u,y) - \tilde{F}'_x(u^*,y)\right|. \end{aligned} \tag{9}$$

From (1) we get that

$$\lim_{n\to\infty}\sup_{0<u<1}\left|n^{\frac{1}{2}}\left(\tilde{F}_{n,1}(u) - u\right) - B_1(u)\right| = 0 \quad a.s.,$$

from which it follows with the Vervaat (1972) lemma that

$$\lim_{n\to\infty}\sup_{0<u<1}\left|n^{\frac{1}{2}}\left(\tilde{F}_{n,1}^{-1}(u) - u\right) + B_1(u)\right| = 0 \quad a.s. \tag{10}$$

Hence, with (10) and (7), the first term of (9) is equal to 0 as $n \to \infty$. With (10) and assumption (B), the second factor of the second term of (9) is going to 0 as $n \to \infty$, and because of the boundedness of B_1, we obtain (8). □

Let $\alpha \in (0, \frac{1}{2})$. For $x \in I_0$ and $y \in \mathbb{R}$, write

$$\begin{aligned} L_x(y) \;:=\; \tfrac{1}{2\alpha}\Big\{ & B_F\big(Q_1(F_1(x)+\alpha), y\big) - B_F\big(Q_1(F_1(x)-\alpha), y\big) \\ & + \tilde{F}'_x\big(F_1(x)+\alpha, y\big)\,\left[B_F(x,\infty) - B_F\left(Q_1\left(F_1(x)+\alpha\right),\infty\right)\right] \\ & - \tilde{F}'_x\left(F_1(x)-\alpha, y\right)\,\left[B_F(x,\infty) - B_F\left(Q_1\left(F_1(x)-\alpha\right),\infty\right)\right]\Big\}. \end{aligned}$$

Note that

$$\sup_{(x,y)\in I_0\times\mathbb{R}}|L_x(y)| < \infty, \tag{11}$$

and that the functions $\{L_x : x \in I_0\}$ are uniformly equicontinuous.

Lemma 2 *Let $\alpha \in (0, \frac{1}{2})$. Under the assumptions (A) and (B), on the probability space of (1),*

$$\lim_{n\to\infty}\sup_{(x,y)\in I_n\times\mathbb{R}}\left|n^{\frac{1}{2}}\left[G_{n,x}(y) - G_x(y)\right] - L_x(y)\right| = 0 \quad a.s.$$

Proof First, rewrite $2\alpha n^{\frac{1}{2}}\left[G_{n,x}(y)-G_x(y)\right]$ in the following way:

$$
\begin{aligned}
(12)\qquad & n^{\frac{1}{2}}\Bigg[\tilde{F}_n\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x)+\alpha\right),y\right)-\tilde{F}\left(F_1(x)+\alpha,y\right)\\
& \qquad -\tilde{F}_n\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x)-\tfrac{\lceil n\alpha\rceil}{n}\right),y\right)+\tilde{F}\left(F_1(x)-\alpha,y\right)\Bigg]\\
(12a)\qquad & = n^{\frac{1}{2}}\left[\tilde{F}_n\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x)+\alpha\right),y\right)-\tilde{F}\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x)+\alpha\right),y\right)\right]\\
(12b)\qquad & + n^{\frac{1}{2}}\left[\tilde{F}\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x)+\alpha\right),y\right)-\tilde{F}\left(F_{n,1}(x)+\alpha,y\right)\right]\\
(12c)\qquad & + n^{\frac{1}{2}}\left[\tilde{F}\left(F_{n,1}(x)+\alpha,y\right)-\tilde{F}\left(F_1(x)+\alpha,y\right)\right]\\
(12d)\qquad & - n^{\frac{1}{2}}\left[\tilde{F}_n\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x)-\tfrac{\lceil n\alpha\rceil}{n}\right),y\right)-\tilde{F}\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x)-\tfrac{\lceil n\alpha\rceil}{n}\right),y\right)\right]\\
(12e)\qquad & - n^{\frac{1}{2}}\left[\tilde{F}\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x)-\tfrac{\lceil n\alpha\rceil}{n}\right),y\right)-\tilde{F}\left(F_{n,1}(x)-\tfrac{\lceil n\alpha\rceil}{n},y\right)\right]\\
(12f)\qquad & - n^{\frac{1}{2}}\left[\tilde{F}\left(F_{n,1}(x)-\tfrac{\lceil n\alpha\rceil}{n},y\right)-\tilde{F}\left(F_1(x)-\alpha,y\right)\right].
\end{aligned}
$$

From (1)

$$
\lim_{n\to\infty}\sup_{(x,y)\in I_n\times\mathbb{R}}\Bigg|n^{\frac{1}{2}}\left[\tilde{F}_n\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x)+\alpha\right),y\right)-\tilde{F}\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x)+\alpha\right),y\right)\right]
-B_F\left(Q_{n,1}\left(F_{n,1}(x)+\alpha\right),y\right)\Bigg|=0\quad a.s.
$$

Because of the uniform continuity of $\tilde{F}$, with (10), (12a) converges almost surely for $n\to\infty$, uniformly in $x\in I_n$ and $y\in\mathbb{R}$, to $B_F\left(Q_1\left(F_1(x)+\alpha\right),y\right)$. Similarly, the expression in (12d) converges almost surely for $n\to\infty$, uniformly in $x\in I_n$ and $y\in\mathbb{R}$, to $-B_F\left(Q_1\left(F_1(x)-\alpha\right),y\right)$.

From Lemma 1, assumption (B), and the uniform continuity of B_1, we have that the expression in (12b) converges uniformly on $I_n\times\mathbb{R}$, as $n\to\infty$, to $-\tilde{F}'_x\left(F_1(x)+\alpha,y\right)\ B_1\left(F_1(x)+\alpha\right)$ almost surely; similarly the expression in (12e) converges to $\tilde{F}'_x\left(F_1(x)-\alpha,y\right)\ B_1\left(F_1(x)-\alpha\right)$ almost surely.

For the convergence of (12c) and (12f), a similar argument as for (12b) and (12e) holds. For the expression in (12f), with (1), (7) and the mean-value theorem,

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{(x,y)\in I_n\times\mathbb{R}} \Bigg| n^{\frac{1}{2}} \left[\tilde{F}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha\rceil}{n}, y\right) - \tilde{F}\left(F_1(x)-\alpha, y\right)\right] \\
&\qquad - \tilde{F}'_x\left(F_1(x)-\alpha, y\right)\, B_F(x,\infty) \Bigg| \\
\leq\ & \lim_{n\to\infty} \sup_{(x,y)\in I_n\times\mathbb{R}} \left| \frac{\tilde{F}\left(F_{n,1}(x) - \frac{\lceil n\alpha\rceil}{n}, y\right) - \tilde{F}\left(F_1(x)-\alpha, y\right)}{F_{n,1}(x) - F_1(x) + \alpha - \frac{\lceil n\alpha\rceil}{n}} - \tilde{F}'_x\left(F_1(x)-\alpha, y\right)\right| \\
&\quad \cdot \lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left| n^{\frac{1}{2}}\left(F_{n,1}(x) - F_1(x)\right)\right| + \sup_{(x,y)\in I_0\times\mathbb{R}} \left|\tilde{F}'_x\left(F_1(x)-\alpha, y\right)\right| \\
&\quad \cdot \lim_{n\to\infty} \sup_{x\in I_n} \left| n^{\frac{1}{2}}\left(F_{n,1}(x) - F_1(x)\right) - B_F(x,\infty)\right| \ =\ 0 \ \ a.s.,
\end{aligned}$$

because of assumption (B). Similarly,

$$\begin{aligned}
\lim_{n\to\infty} \sup_{(x,y)\in I_n\times\mathbb{R}} \Bigg| & n^{\frac{1}{2}} \left[\tilde{F}\left(F_{n,1}(x)+\alpha, y\right) - \tilde{F}\left(F_1(x)+\alpha, y\right)\right] \\
& - \tilde{F}'_x(F_1(x)+\alpha, y)\, B_F(x,\infty) \Bigg| \ =\ 0 \quad a.s. \qquad \square
\end{aligned}$$

Lemma 3 *Under (A) and (B), we have for fixed $s \in (0,1)$, on the probability space of (1),*

$$\lim_{n\to\infty} \sup_{x\in I_n} \left|Q_{n,x}(s) - Q_x(s)\right| \ =\ 0 \quad a.s. \tag{13}$$

and

$$\lim_{n\to\infty} \sup_{x\in I_n} \left| n^{\frac{1}{2}}\left(Q_{n,x}(s) - Q_x(s)\right) \ +\ Q'_x(s) L_x\left(Q_x(s)\right)\right| \ =\ 0 \quad a.s. \tag{14}$$

Proof Since $G_x(\cdot)$ is increasing on $\mathbb{R}$,

$$\begin{aligned}
\left|Q_{n,x}(s) - Q_x(s)\right| &= \left|Q_x\left(G_x\left(Q_{n,x}(s)\right)\right) - Q_x(s)\right| \\
&= \left| \frac{Q_x\left(G_x\left(Q_{n,x}(s)\right)\right) - Q_x(s)}{G_x\left(Q_{n,x}(s)\right) - s} \left(G_x\left(Q_{n,x}(s)\right) - s\right)\right|. \qquad (15)
\end{aligned}$$

Because Q_x is differentiable, the mean-value theorem yields that the right-hand side of (15) is equal to

$$\left|Q'_x(s^*_n)\left(G_x\left(Q_{n,x}(s)\right) - s\right)\right| \ =\ \left| \frac{1}{G'_x\left(Q_x(s^*_n)\right)} \left(G_x\left(Q_{n,x}(s)\right) - s\right)\right|, \tag{16}$$

for some s_n^* between $G_x\left(Q_{n,x}(s)\right)$ and s. From Lemma 2 it follows that

$$\begin{aligned}
&\limsup_{n\to\infty}\sup_{x\in I_n}\left|n^{\frac{1}{2}}\left(G_x\left(Q_{n,x}(s)\right)-s\right)+L_x\left(Q_{n,x}(s)\right)\right| \\
&= \lim_{n\to\infty}\sup_{x\in I_n}\left|-n^{\frac{1}{2}}\left(G_{n,x}\left(Q_{n,x}(s)\right)-G_x\left(Q_{n,x}(s)\right)\right)+L_x\left(Q_{n,x}(s)\right)\right| \;=\; 0 \;\; a.s., \qquad (17)
\end{aligned}$$

and hence, since L_x bounded,

$$\lim_{n\to\infty}\sup_{x\in I_n}\left|G_x\left(Q_{n,x}(s)\right)-s\right| = 0 \quad a.s. \qquad (18)$$

Note that $\inf_{x\in I_0} Q_1\left(F_1(x)+\alpha\right)-Q_1\left(F_1(x)-\alpha\right)>0$ and, for $0<s_1<s<s_2<1$, uniformly in $x\in I_0$ almost surely for n large enough, $Q_x(s_n^*)\in\left[Q_2(2\alpha s_1), Q_2\left(1-2\alpha(1-s_2)\right)\right]$. Hence

$$\inf_{x\in I_n} G_x'\left(Q_x(s_n^*)\right) \;\geq\; \inf_{x\in I_0}\frac{1}{2\alpha}\int_{Q_1(F_1(x)-\alpha)}^{Q_1(F_1(x)+\alpha)} f\left(u, Q_x(s_n^*)\right)du \;>\; 0. \qquad (19)$$

This proves (13).

From (15)–(17), with s_n^* between $G_x\left(Q_{n,x}(s)\right)$ and s, it follows immediately that the left-hand side of (14) is equal to

$$\begin{aligned}
&\limsup_{n\to\infty}\sup_{x\in I_n}\left|-Q_x'(s_n^*)L_x\left(Q_{n,x}(s)\right)+Q_x'(s)L_x\left(Q_x(s)\right)\right| \\
&= \limsup_{n\to\infty}\sup_{x\in I_n}\left|-Q_x'(s_n^*)\left[L_x\left(Q_{n,x}(s)\right)-L_x\left(Q_x(s)\right)\right]-L_x\left(Q_x(s)\right)\left[Q_x'(s_n^*)-Q_x'(s)\right]\right| . \qquad (20)
\end{aligned}$$

Since the functions $\{L_x : x\in I_0\}$ are uniformly equicontinuous, with (13),

$$\lim_{n\to\infty}\sup_{x\in I_n}\left|L_x\left(Q_{n,x}(s)\right)-L_x\left(Q_x(s)\right)\right| = 0 \quad a.s. \qquad (21)$$

As in (19), we have $\inf_{x\in I_n} G_x'\left(Q_x(s)\right)>0$. With assumption (B), (13), and (19), it follows that

$$\begin{aligned}
&\limsup_{n\to\infty}\sup_{x\in I_n}\left|Q_x'(s_n^*)-Q_x'(s)\right| \\
&= \lim_{n\to\infty}\sup_{x\in I_n}\left|\frac{\tilde{F}_y'\left(F_1(x)+\alpha, Q_x(s)\right)-\tilde{F}_y'\left(F_1(x)+\alpha, Q_x(s_n^*)\right)}{2\alpha G_x'\left(Q_x(s_n^*)\right)G_x'\left(Q_x(s)\right)}\right. \\
&\qquad \left.+\frac{\tilde{F}_y'\left(F_1(x)-\alpha, Q_x(s_n^*)\right)-\tilde{F}_y'\left(F_1(x)-\alpha, Q_x(s)\right)}{2\alpha G_x'\left(Q_x(s_n^*)\right)G_x'\left(Q_x(s)\right)}\right| \;=\; 0 \quad a.s. \qquad (22)
\end{aligned}$$

Equations (11), (19), (21) and (22) give that (20) is equal to zero and hence equation (14) is proven. □

5.2 Proof of Theorem 1

For the proof of Theorem 1 it suffices to show that for $s \in \left\{\frac{1}{4}, \frac{3}{4}\right\}$, on the probability space of (1),

$$\sup_{x \in I_n} \Bigg| \, n^{\frac{1}{2}} \, \Bigg\{ \Big[F_n\left(x, Q_{n,x}\left(s\right)\right) - F_n\left(Q_{n,1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}\right), Q_{n,x}(s)\right)\Big] \\ - \big[F\big(x, Q_x\left(s\right)\big) - F\big(Q_1\left(F_1(x) - \alpha\right), Q_x(s)\big)\big] \Bigg\} \\ - B_s(x) \Bigg| \to 0 \quad a.s. \;\; as \;\; n \to \infty. \tag{23}$$

First, rewrite in the following way:

$$n^{\frac{1}{2}} \, \Bigg\{ \Big[F_n\big(x, Q_{n,x}\left(s\right)\big) - F_n\left(Q_{n,1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}\right), Q_{n,x}(s)\right)\Big] \\ - \Big[F\big(x, Q_x\left(s\right)\big) - F\big(Q_1\left(F_1(x) - \alpha\right), Q_x(s)\big)\Big] \Bigg\} \tag{24}$$

$$= \; n^{\frac{1}{2}} \; \big\{F_n\big(x, Q_{n,x}\left(s\right)\big) - F\big(x, Q_{n,x}\left(s\right)\big)\big\} \tag{24a}$$

$$+ \; n^{\frac{1}{2}} \; \big\{F\big(x, Q_{n,x}\left(s\right)\big) - F\big(x, Q_x\left(s\right)\big)\big\} \tag{24b}$$

$$- \; n^{\frac{1}{2}} \; \Big\{\tilde{F}_n\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}\right), Q_{n,x}(s)\right) - \tilde{F}\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}\right), Q_{n,x}(s)\right)\Big\} \tag{24c}$$

$$- \; n^{\frac{1}{2}} \; \Big\{\tilde{F}\left(\tilde{F}_{n,1}^{-1}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}\right), Q_{n,x}(s)\right) - \tilde{F}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}, Q_{n,x}(s)\right)\Big\} \tag{24d}$$

$$- \; n^{\frac{1}{2}} \; \Big\{\tilde{F}\left(F_{n,1}(x) - \tfrac{\lceil n\alpha \rceil}{n}, Q_{n,x}(s)\right) - \tilde{F}\left(F_1(x) - \alpha, Q_{n,x}(s)\right)\Big\} \tag{24e}$$

$$- \; n^{\frac{1}{2}} \; \big\{\tilde{F}\left(F_1(x) - \alpha, Q_{n,x}(s)\right) - \tilde{F}\left(F_1(x) - \alpha, Q_x(s)\right)\big\}. \tag{24f}$$

The subtrahends (24c), (24d) and (24e) are equal to (12d), (12e) and (12f), respectively, evaluated at $y = Q_{n,x}(s)$. Therefore, with Lemma 3, uniformly in $x \in I_n$ for $n \to \infty$, with a similar reasoning as for (12d), (24c) converges to $-B_F\left(Q_1\left(F_1(x) - \alpha\right), Q_x(s)\right)$ almost surely. Analogous to (24c), (24a) converges uniformly in $x \in I_n$ for $n \to \infty$ almost surely to $B_F\left(x, Q_x(s)\right)$. Taking moreover (B) into account, it follows immediately from Lemma 1 and Lemma 3 that (24d) converges to $\tilde{F}'_x\left(F_1(x) - \alpha, Q_x(s)\right) \; B_1\left(F_1(x) - \alpha\right)$ almost surely and (24e) converges to $-\tilde{F}'_x\left(F_1(x) - \alpha, Q_x(s)\right) \; B_F\left(x, \infty\right)$ almost surely.

For (24b), we have

$$\limsup_{n \to \infty} \sup_{x \in I_n} \Big| n^{\frac{1}{2}} \left[F\left(x, Q_{n,x}\left(s\right)\right) - F\left(x, Q_x\left(s\right)\right)\right] \; + \; F'_y\left(x, Q_x(s)\right) Q'_x(s) L_x\left(Q_x(s)\right)\Big|$$

$$
\begin{aligned}
&= \limsup_{n\to\infty} \sup_{x\in I_n} \left| \frac{F\left(x, Q_{n,x}(s)\right) - F\left(x, Q_x(s)\right)}{Q_{n,x}(s) - Q_x(s)}\; n^{\frac{1}{2}} \left(Q_{n,x}(s) - Q_x(s)\right) \right. \\
&\quad \left. + F'_y\left(x, Q_x(s)\right) Q'_x(s) L_x\left(Q_x(s)\right) \right| \\
&\leq \limsup_{n\to\infty} \sup_{x\in I_n} \left| \frac{F\left(x, Q_{n,x}(s)\right) - F\left(x, Q_x(s)\right)}{Q_{n,x}(s) - Q_x(s)} \left[n^{\frac{1}{2}} \left(Q_{n,x}(s) - Q_x(s)\right) + Q'_x(s) L_x\left(Q_x(s)\right) \right] \right| \\
&\quad + \limsup_{n\to\infty} \sup_{x\in I_n} \left| Q'_x(s) L_x\left(Q_x(s)\right) \left[\frac{F\left(x, Q_{n,x}(s)\right) - F\left(x, Q_x(s)\right)}{Q_{n,x}(s) - Q_x(s)} - F'_y\left(x, Q_x(s)\right) \right] \right| . \qquad (25)
\end{aligned}
$$

With assumption (B), Lemma 3, and the mean-value theorem, it follows that

$$
\lim_{n\to\infty} \sup_{x\in I_n} \left| \frac{F\left(x, Q_{n,x}(s)\right) - F\left(x, Q_x(s)\right)}{Q_{n,x}(s) - Q_x(s)} - F'_y\left(x, Q_x(s)\right) \right| = 0 \quad a.s., \qquad (26)
$$

and with assumption (A) we get

$$
\sup_{x\in I_n} F'_y\left(x, Q_x(s)\right) \leq \sup_{y\in\mathbb{R}} f_2(y) < \infty,
$$

which, with (26), bounds the first factor of the first term of the right-hand side of (25). Lemma 3 yields that the second factor of this term is equal to zero almost surely. From (11), (19), and (26) it now follows directly that (25) is equal to zero almost surely.

Accordingly (24f) converges uniformly in $x \in I_n$ for $n \to \infty$, almost surely to

$$
F'_y\left(Q_1\left(F_1(x) - \alpha\right), Q_x(s)\right)\; Q'_x(s)\; L_x\left(Q_x(s)\right) \; . \qquad \square
$$

Acknowledgements

We thank Irène Gijbels for kindly providing the Prague temperature data and Günther Sawitzki for stimulating discussions.

References

Antonidiadis, A., and Gijbels, I. (2002), "Detecting Abrupt Changes by Wavelet Methods," *Journal of Nonparametric Statistics*, 14, 7–29.

Carlstein, E. (1988), "Nonparametric Change-Point Estimation," *The Annals of Statistics*, 16, 188–197.

Cobb, G.W. (1978), "The Problem of the Nile: Conditional Solution to a Change-Point Problem," *Biometrika*, 64, 243–251.

Dempfle, A., and Stute, W. (2002), "Nonparametric Estimation of a Discontinuity in Regression," *Statistica Neerlandica*, 56, 233–242.

Gijbels, I., and Goderniaux, A.-C. (2004), "Bandwidth Selection for Changepoint Estimation in Nonparametric Regression," *Technometrics*, 46, 76–86.

Gijbels, I., Hall, P. and Kneip, A. (1999), "On the Estimation of Jump Points in Smooth Curves," *Annals of the Institute of Statistical Mathematics*, 51, 231–251.

Gijbels, I., Lambert, A., and Qiu, P. (2007), "Jump-Preserving Regression and Smoothing Using Local Linear Fitting: a Compromise," *Annals of the Institute of Statistical Mathematics*, 59, 235–272.

Grégoire, G., and Hamrouni, Z. (2002), "Change Point Estimation by Local Linear Smoothing," *Journal of Multivariate Analysis*, 83, 56–83.

Hall, P., and Titterington, D.M. (1992), "Edge-Preserving and Peak-Preserving Smoothing," *Technometrics*, 34, 429–440.

Horváth, L., and Kokoszka, P. (2002), "Change-Point detection With Nonparametric Regression," *Statistics*, 36, 9–31.

Horváth, L., Kokoszka, P., and Steinebach, J. (1999), "Testing for Changes in Multivariate Dependent Observations With an Application to Temperature Changes," *Journal of Multivariate Analysis*, 68, 96–119.

Kim, C.S., and Marron, J.S. (2006), "SiZer for Jump Detection," *Journal of Nonparametric Statistics*, 18, 13–20.

Kraus, E.B. (1955), "Secular Changes of Tropical Rainfall Regimes," *Quarterly Journal of the Royal Meteorological Society*, 81, 198–210.

Loader, C.R. (1996), "Change Point Estimation Using Nonparametric Regression," *The Annals of Statistics*, 24, 1667–1678.

McDonald, J.A., and Owen, A.B. (1986), "Smoothing with Split Linear Fits," *Technometrics*, 28, 195–208.

Müller, H.-G. (1992), "Change-Points in Nonparametric Regression Analysis," *The Annals of Statistics*, 20, 737–761.

Park, C., and Kim, W.-C. (2006), "Wavelet Estimation of a Regression Function With a Sharp Change Point in a Random Design," *Journal of Statistical Planning and Inference*, 136, 2381–2394.

Sánchez-Borrego, I.R., Martínez-Miranda, M.D., and González-Carmona, A. (2006), "Local Linear Kernel Estimation of the Discontinuous Regression Function," *Computational Statistics*, 21, 557–569.

Vervaat, W. (1972), "Functional Central Limit Theorems for Processes with Positive Drift and Their Inverses," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 23, 245–253.

Wang, Y. (1995), "Jump and Sharp Cusp Detection by Wavelets," *Biometrika*, 82, 385–397.

Wu, J.S., and Chu, C.K. (1993), "Kernel-Type Estimation of Jump Points and Values of a Regression Function," *The Annals of Statistics*, 21, 1545–1566.